Effect of Correlation on Signal Detection in Arctic Under-Ice Noise *

P.A. Nielsen J.B. Thomas

Department of Electrical Engineering
Princeton University
Princeton, N.J. 08544

Abstract

Signal detection in a large segment of non-Gaussian and non-stationary Arctic under-ice noise, which contains both high power narrow-band and impulsive components, is examined. It is shown that the correlation functions of sub-segments of data change significantly, and if ignored, can degrade the performance of a detector. For a false alarm probability of 0.05 and a known constant signal, the matched filter was on average 17.6% better than a detector designed assuming independent noise samples. It is also shown that pre-processing the data with an adaptive notch filter, then using the matched filter will result in a further improvement of about 6%. Additionally, the effect two different signal shapes have on the performance of the matched filter is examined.

Introduction

With the goal of creating a database of under-ice noise, in 1980 a multi-institutional research program collaborated to record Arctic under-ice data. The data, FRAM II, was recorded on 23-24 April 1980, from a pack ice floe, at 86°N 25°W. An omnidirectional hydrophone, radio linked to a receiver, was suspended to a depth of 91m in 4000m deep water. The data, recorded on an analog device, was then bandpass filtered from 0.01-5kHz. The particular data segment being analyzed was recorded on the 23 April from 11:30 to 11:40 pm. Subsequent to the bandpass filter previously mentioned, the data was passed through a low pass filter with cutoff of 2.5 kHz, and digitized with a sampling frequency of 10 kHz. A total of 6,150,144 data samples were taken, which conveniently breaks down to 6006 records with each record containing exactly 1024 samples [1,2].

Several authors have looked at this particular data segment in varying degrees of detail. Dwyer [1] took several different approaches to analyzing the under-ice data. He examined the first four central moments in the time domain and concluded that the noise is non-stationary and non-Gaussian. A similar analysis carried out in the frequency domain yielded comparable results. The energy spectrum indicated that the ice noise was heavier tailed than a Gaussian source processed in the same fashion, in agreement with the results of Greene and Buck [3]. From an examination of the spectrum and the time domain plots of the data, both narrow-band and impulsive components were identified. The existence of these components was not surprising, since the floe was moving slowly throughout the experiment. As suggested by the work of Milne and Ganton [4], the movement of the floe would cause rifting and cracking of the ice.

Veitch and Wilks [2] took a slightly different approach since their objective was to develop a statistical model for the under-ice noise. Their work confirms the presence of narrow-band and impulsive components, but suggests that the background noise is Gaussian. By examining the first four central moments of the data in detail, they noted a strong correlation between records with large kurtosis and large skew. They also determined that a large kurtosis was associated with high amplitude bursts. In addition, it was proposed that the Arctic data could be modeled in the frequency domain as a mixture of a Gaussian background process, a sum of a random number of narrow-band components, each with random phase and amplitude, and what is, in some sense, an error term. The main drawbacks of the model are that the impulsive component, which has been verified as being present, is not modeled specifically. Also the weights and number of sinusoidal components are not modeled constructively; that is, for some arbitrary segment of the data, the number of tones is unknown.

Recently, Nielsen and Thomas [5] proposed that the univariate statistics of the Arctic data could be modeled by the generalized Gaussian family, with the parameters of the model estimated from the data. The model was shown to be a reasonably accurate representation of the empirical probability mass functions for blocks (5 records) of data, but was not accurate enough for the resulting Neyman-Pearson optimal detector to show an improvement over a simple linear structure.

Time Correlation

Let a sample function of a continuous parameter random process be sampled at the Nyquist frequency, yielding samples x_i. Define the time correlation functions as

$$P_k = \frac{1}{N} \sum_{i=1}^{N} x_i x_{i+k}$$

Equivalently, for any realization $\underline{x}$, the Fourier transform of the realization can be used to compute the power spectrum, which can then be inverse Fourier transformed to obtain an estimate of the time correlation function. The Arctic data was sampled at 10 kHz, which is twice the Nyquist frequency. For these computations the data was decimated by a factor of two so that the sampling frequency was equal to the Nyquist frequency; thus, the correlation functions were

* Presented at the *22nd Annual Asilomar Conference on Signals, Systems, and Computers*, Asilomar Ca., Oct. 31 - Nov. 2, 1988.

computed using N=512.

Since time correlation functions for real processes are even about the origin. k=0, the graphs of typical correlation functions, shown in Figures 1 - 4, are plotted only for k=0, 1, 2,..., 256. In general, it was found that the correlation function took on one of four different shapes. The time correlation function of record 0901, Figure 1, shows a mixture of a decaying low frequency component with a small amplitude high frequency component. The main variations for this type of function are the rate of decay and the amplitude of the high frequency component.

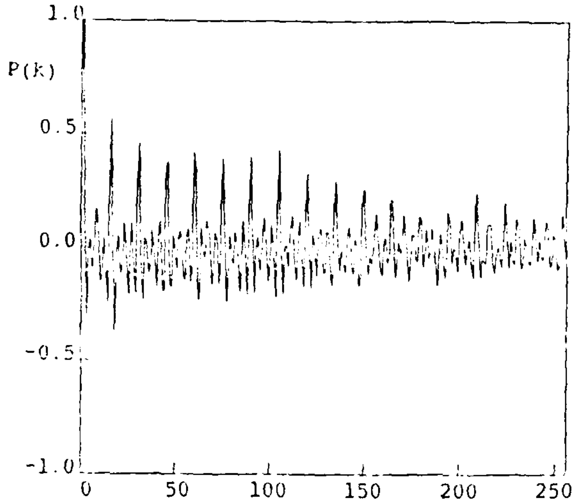

Figure 1 - Time Correlation Record 0901

Figure 2 shows the correlation function for record 0905, approximately one-half second later than record 0901. The correlation drops rapidly as k increases, and then has small amplitude variations about zero. A close examination of this correlation function shows that significant correlations exist for the first three shifts. The drastic change in

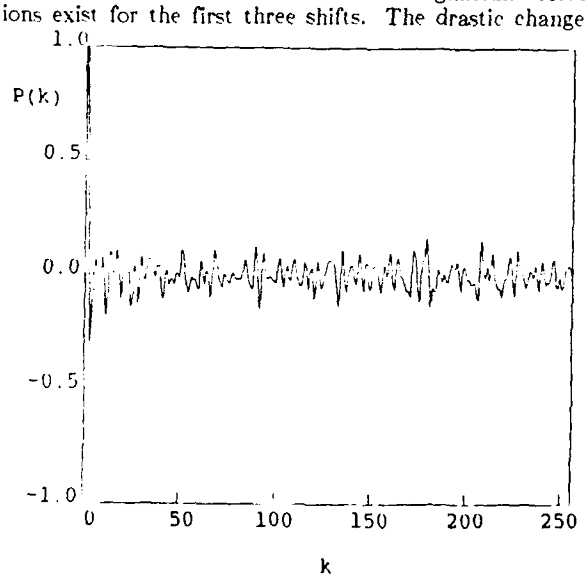

Figure 2 - Time Correlation Record 0905

the shape of the correlation function in a short period of time, as illustrated by Figures 1 and 2, is not unusual for this data.

The third typical correlation function is shown in Figure 3. For record 1205 the correlation is a mixture of a low frequency and a high frequency component. The result is that the correlation function looks like a corrupted AM signal. The principal variations with the general shape are in the magnitude of the correlation for small k, and the decay rate of the high frequency component.

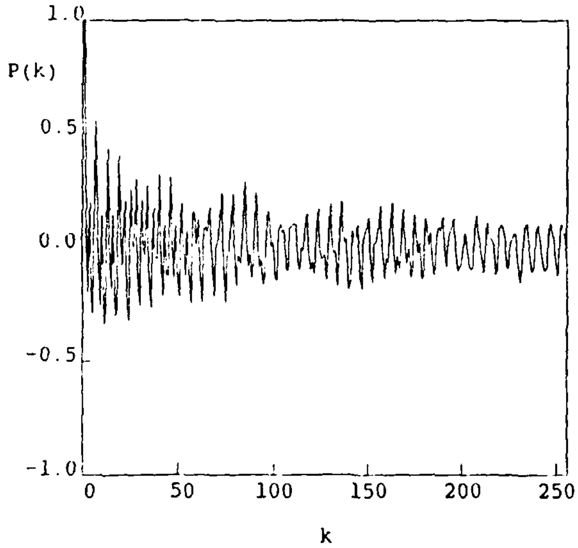

Figure 3 - Time Correlation Record 1205

One of the most interesting correlation functions belongs to the final category. In Figure 4 the correlation function for data record 3905 is shown. The correlation for this data is very high, and periodic, with peak correlations greater than 0.5 for shifts greater than zero. The peak correlation for shifts other than zero were found to be plus or minus one for some data records, and the envelope shape varied greatly. The main factor correlation functions of this type have in common is the presence of a single, high powered periodic component.

Adaptive Notch Filter

The correlation between data samples, which has been shown to be significant in the Arctic under-ice noise, provides a complicating factor in the problem of signal detection. This results primarily from inability to model the multivariate density if the data is known to be non-Gaussian. A common approach to working with highly correlated noise is to pre-whiten the data, ideally resulting in independent noise samples. Sub-optimal pre-whiteners, which are designed to reduce instead of eliminate the correlation between data samples, are used with real noise sources.

For environments such as the Arctic data, which contain a combination of narrow-band and wideband processes, an adaptive line enhancer [6] is commonly used to separate the two processes. Suppose the received vector Y is the sum of two components; a white background component N and a sinusoidal component S. We can represent this mathematically as

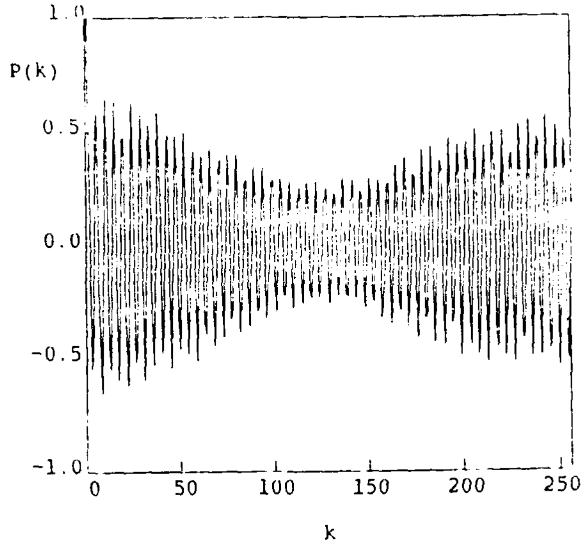

Figure 4 - Time Correlation Record 3905

$$Y = N + S$$

In Figure 5 we design the filter to be an adaptive notch filter (ANF) with a gain of unity everywhere except at a specified band of frequencies, where the gain is zero. The

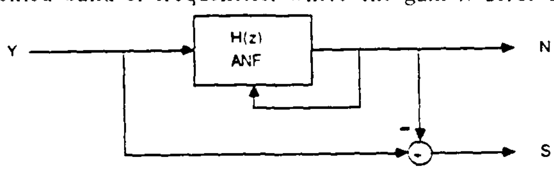

Figure 5 - Adaptive Line Enhancer

filter output will be the white noise N and the output of the difference operator will be the sinusoid S. We define the ANF as follows [7]

$$H(z) = \frac{z^2 + bz + 1}{z^2 + \gamma bz + \gamma^2}$$

Then, we have

$$\text{bandwidth} \approx 2(1-\gamma), \qquad \gamma \in [0,1]$$

$$\text{normalized center frequency} \approx \cos^{-1}(-b/2), \qquad b \in [-2,2]$$

For an input sequence y_k and output sequence of the ANF u_k, we have

$$u_k = y_k + by_{k-1} + y_{k-2} - b\gamma u_{k-1} - \gamma^2 u_{k-2}$$

For a fixed notch width, the only adaptive parameter of the filter is b, which controls the center frequency of the notch. To derive the update equation for b, we minimize u_k^2 using a gradient search algorithm, then

$$b_{k+1} = b_k + step\ u_k\,(y_{k-1} - \gamma\, u_{k-1})$$

The variable *step*, which can be either fixed or adaptively set, controls the adaptation speed of the filter and the variance of the filter center frequency once steady state has been achieved. That is, a value of *step* too small will result in a long time for the filter to adjust to the proper center frequency, however, once there, only small changes in the filter center frequency will occur. The converse results will hold for a value of *step* too large. One advantage of the ANF is that it can be implemented easily as a time domain processer, in contrast with the frequency domain preprocessor proposed by Dwyer [8].

As an example of the effect the ANF can have on the correlation function, Figures 6 - 7 show the correlation function for record 4201 before and after filtering. The ANF essentially removes the high frequency component which dominates the correlation function, resulting in data which is much less correlated. The results are not always as dramatic as with record 4201. Specifically, if the data contains no dominant narrow-band component, little difference in the correlation function after filtering is observed. Also, when the data contains multiple narrow-band components, the ANF only removes the highest powered; thus the filtered data still has a correlation function with periodic components.

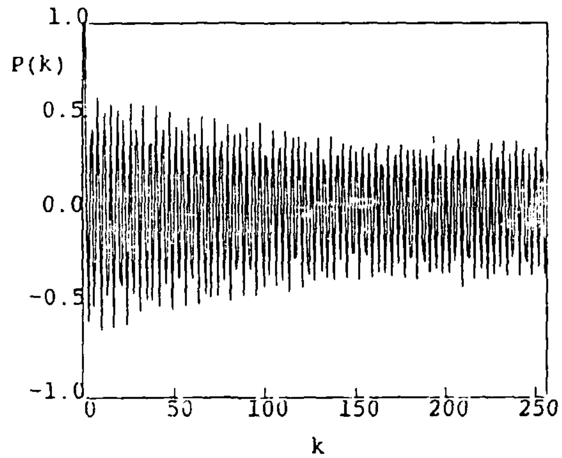

Figure 6 - Time Correlation Record 4201

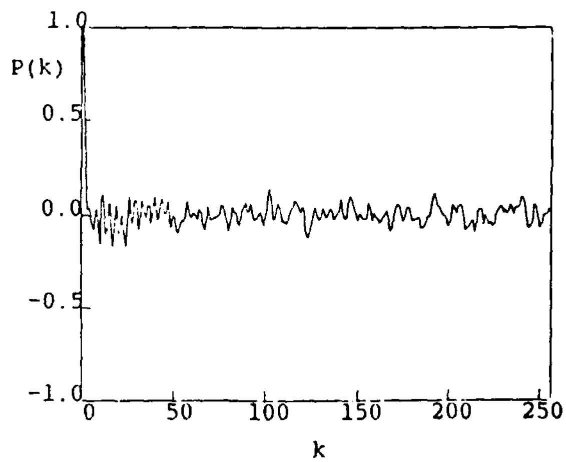

Figure 7 - Time Correlation Filtered Record 4201

Signal Detection

Consider the following simple detection problem. Let M be the number of samples taken per decision, and let $\underline{N}$ be the M-vector of samples from the zero mean noise process and $\underline{s}$ be the known signal M-vector. Let f_M represent the M-variate density of the noise. Define a binary hypothesis problem in the following manner

$$H_0: \quad \underline{X} = \underline{N}$$
$$H_1: \quad \underline{X} = \underline{N} + \underline{s}$$

where $\underline{X}$ is the vector received by the detector and is of length M. Let

$$\alpha = \textit{Probability (decide } H_1 \mid H_0 \textit{ true)}$$
$$\beta = \textit{Probability (decide } H_1 \mid H_1 \textit{ true)}$$

The Neyman-Pearson criterion can be written as follows. Fix some α_0. Define the optimal detector to be the detector which maximizes the power β for any $\alpha \leq \alpha_0$. Define the likelihood ratio l as follows

$$l(\underline{X}) = \log \frac{f_M(\underline{X} - \underline{s})}{f_M(\underline{X})}$$

Then the optimal detector can be written as

$$l(\underline{x}) \geq T \longrightarrow H_1$$
$$l(\underline{x}) < T \longrightarrow H_0$$

The threshold T is chosen to satisfy the constraints on the power β and on the level α [9].

As mentioned before, for non-Gaussian data which is correlated, estimating the multivariate density of the noise may not be possible. One approach taken to circumvent this problem is to assume that the multivariate statistics are Gaussian; then using the Neyman-Pearson criteria, the optimal likelihood ratio is the matched filter. If R is the $M \times M$ covariance matrix of the noise, then the matched filter is defined as

$$l(\underline{x}) = \underline{s}^T R^{-1} \underline{x}$$

For independent noise samples, $EN_iN_j = \delta_{ij}\sigma^2$ and the covariance matrix $R = \sigma^2 I$, where I is an $M \times M$ identity matrix. Consequently, the likelihood ratio reduces to $l(\underline{x}) = \underline{s}^T \underline{x}/\sigma^2$. The resulting detector is then a correlator. In addition only the variance of the noise σ^2 need be estimated.

For correlated data the estimation of the elements of the covariance matrix is necessary. While the Arctic noise does not appear to be stationary in any sense, we will assume that within a record (that is 1024 contiguous samples) the data is at least wide sense stationary and ergodic, and that the mean and variance of the noise are known exactly. Under the assumption of ergodicity, the correlation function is equivalent to a vector of expected values, $P_k = EX_jX_{j+k}$. Define the covariance matrix of the noise as

$$R_{ij} = EX_iX_j = R_{ji} = R_{i-j} = P_{i-j}$$

Each record contains N=1024 samples; thus P_k is computed for k=0, 1, ..., 1023. For a signal length of M, only *the first M correlations are used in the formation of R*. The resulting covariance matrix R is symmetric and Toeplitz. For R to be a valid invertible covariance matrix, we require that it be positive definite.

For convenience, the data was normalized such that $P_0 = 1$. Thus, the covariance matrix is $\sigma^2 R$, where σ^2 is the variance of the record and R is the normalized covariance matrix. Henceforth, we refer to the normalized covariance matrix as the covariance matrix. A signal length of $M = 4$ was used for all simulations. For the entire data set, 6006 covariance matrices were computed. The minimum eigenvalue was always strictly greater than zero, insuring that R was indeed positive definite, and typical values for the four eigenvalues were 0.0077, 0.163, 1.216, and 2.581. In addition, a comparison of covariance matrices showed that the elements changed slowly over time. This suggests the assumptions of wide sense stationarity and ergodicity are not unreasonable for such a short segment of data (1 record $\approx$ 0.1 seconds).

To determine the effect that correlation between data samples has on the signal detection problem, the power β for the following detectors was compared when M=4.

[D1] Correlator ($R = \sigma^2 I$)
[D2] Matched Filter
[D3] ANF - Matched Filter

For [D1]-[D3], in order to compensate for the non-stationary characteristic of the noise, parameters were estimated on a record-by-record basis. For [D1], σ^2 was estimated; for [D2]-[D3], the covariance matrix R was estimated, using the method described previously. In particular, for [D3], the ANF was used as a preprocessor, and the covariance matrix was estimated using the filtered data. For convenience the data was grouped into blocks of 20 records. Thus each block contained 20480 data samples, and represented a time interval of about 2 seconds. We had a total of 300 blocks, and the performance was evaluated for each block.

In Table 1, the mean and variance of the ratio of the power levels for [D1] to [D2] are shown, along with the minimum and maximum values of the ratio, and the performance improvement. We defined the performance improvement to be

$$I(Dj, Di) = \left[\frac{\beta_{[Di]} - \beta_{[Dj]}}{\beta_{[Dj]}}\right] \times 100$$

The results for detectors [D2] and [D3] are presented in Table 2. Examining Tables 1 and 2, we se that the largest improvements in performance are for $\alpha = 0.05$; and while the performance improvement decreases as α increases, the improvement is still significant for $\alpha = 0.20$. For $\alpha = 0.05$, [D2] is approximately 17.6% better and [D3] is approximately 29.2% better, on average, than [D1]. For $\alpha = 0.20$, the improvements are smaller, but at 12.4% and 19.3%, they are still significant.

Given that the Arctic data is highly correlated at times, the improvement that *the matched filter has over* the simple correlator is understandable. The additional improvement which occurs when the ANF is used as a preprocessor is not as clear. It is well known that the output of a linear recursive filter is more Gaussian than the input. Thus using the ANF as a pre-processor not only whitens the data, by removing the highest powered periodic component, but also results in data which is more Gaussian. It

Ratio of Power Levels between [D1] and [D2]

β_{D1}/β_{D2}				
α	0.05	0.10	0.15	0.20
mean	0.850	0.862	0.877	0.889
variance	6.99e-03	5.43e-03	5.00e-03	4.28e-03
minimum	0.587	0.630	0.638	0.679
maximum	1.072	0.981	0.992	0.989
I(D1,D2)	17.62	15.99	14.06	12.46

Table 1

Ratio of Power Levels between [D2] and [D3]

β_{D2}/β_{D3}				
α	0.05	0.10	0.15	0.20
mean	0.910	0.926	0.927	0.943
variance	5.82e-03	3.37e-03	2.04e-03	1.49e-03
minimum	0.597	0.639	0.712	0.753
maximum	1.067	1.025	1.013	1.002
I(D2,D3)	9.85	7.99	6.77	6.06

Table 2

is clear from these results, that disregarding the correlation between data samples will result in a performance degradation.

The minimum and maximum values of the ratios are also interesting. While the maximum values of the ratio are near unity, indicating that the detectors have similar performance, the minimum values range from 0.58 to 0.75, indicating that under some conditions one detector is much better than the other. In particular, blocks 80-90 and 170-210 can be shown by an analysis of their correlation functions to contain periodic components. For these cases, [D2] is about 25% better and [D3] is approximately 47% better than [D1] for $\alpha = 0.20$. Figures 8-9 are plots of β_{D1}/β_{D2} and β_{D2}/β_{D3} for $\alpha = 0.20$ to illustrate more clearly the improvements.

One may challenge the validity of using the same data to estimate the covariance matrix and compute the detection probabilities. Tests were run where the covariance matrix was updated every record, every fifth record, every tenth record, and every twentieth record. It was found that the performance levels did not change appreciably for updates occurring as slowly as every ten records. If the update was performed every twenty records, once a block, only a small performance loss occurred.

For all computations, a constant signal was used such that the sum of the elements of the signal vector equaled the noise standard deviation, i.e. input S/N=1. For the matched filter [D2], the output S/N can be written as

$$\frac{S}{N} = \underline{s}^T R^{-1} \underline{s}$$

For any positive definite matrix A such that λ_{min}, λ_{max} are the minimum and maximum eigenvalues of A, by Rayleigh's principle we have [10]

$$\lambda_{min} \leq \frac{\underline{x}^T A \underline{x}}{\underline{x}^T \underline{x}} \leq \lambda_{max}$$

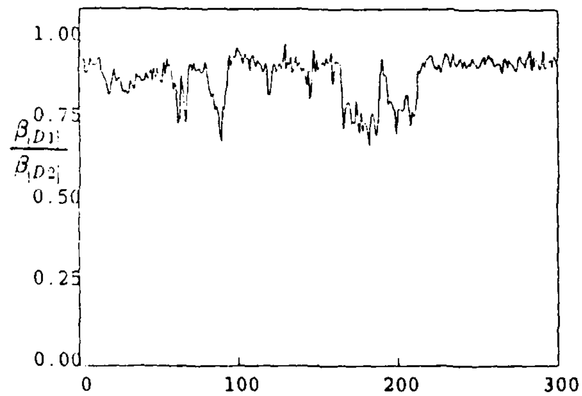

Figure 8 - Ratio of Power Levels between [D1] and [D2] for $\alpha = 0.20$

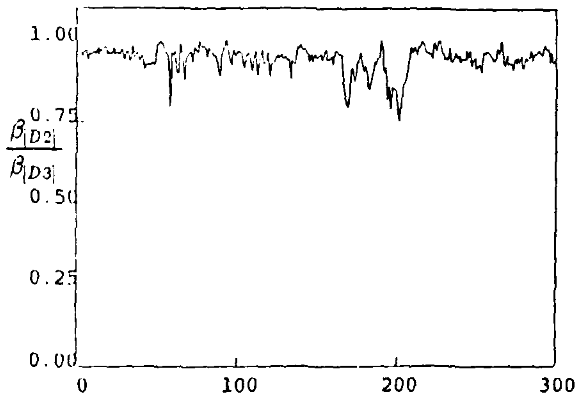

Figure 9 - Ratio of Power Levels between [D2] and [D3] for $\alpha = 0.20$